

AD-A040 995

DEPARTMENT OF DEFENCE CANBERRA (AUSTRALIA) SCIENTIF--ETC F/G 5/2
STORAGE, RETRIEVAL AND DISSEMINATION OF INFORMATION: CURRENT DE--ETC(U)
AUG 76 J CLARKE
STIB-2/76

UNCLASSIFIED

NL

1 OF 1
AD
A040995



END
DATE
FILMED
7-77

ADA 040395

UNCLASSIFIED

11

STIB REPORT 2/76 ✓

AR NO. 000343



DEPARTMENT OF DEFENCE
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
SCIENTIFIC AND TECHNICAL INFORMATION BRANCH ✓
CANBERRA, ACT

STIB REPORT 2/76

STORAGE, RETRIEVAL AND DISSEMINATION OF INFORMATION:
CURRENT DEVELOPMENTS (U)

APPROVED
FOR PUBLIC RELEASE J. CLARKE

DDC
RECEIVED
JUN 28 1977
A

DISTRIBUTION LIMITED TO AUSTRALIAN GOVERNMENT AGENCIES ONLY.
OTHER REQUESTS FOR THIS DOCUMENT MUST BE REFERRED TO THE
SCIENTIFIC AND TECHNICAL INFORMATION BRANCH, DEPARTMENT OF
DEFENCE, CANBERRA.

DDC FILE COPY

COPY NO.

AUGUST 1976

UNCLASSIFIED

Statement A
per US-Aus agreement
RN 6-28-77

THE UNITED STATES NATIONAL
TECHNICAL INFORMATION SERVICE
IS AUTHORIZED TO
REPRODUCE AND SELL THIS REPORT

APPROVED
FOR PUBLIC RELEASE

UNCLASSIFIED

AR No: AR 000343

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
SCIENTIFIC AND TECHNICAL INFORMATION BRANCH

14

STIB REPORT-2/76

6

STORAGE, RETRIEVAL AND DISSEMINATION OF INFORMATION:

CURRENT DEVELOPMENTS, (U)

10

J./Clarke

11

Aug 76

This paper provides a

S U M M A R Y

Australian

A literature survey of current developments in storage, retrieval and dissemination of information and associated technologies, with assessment of their application to the Department of Defence. 780 refs.

12 35p.

Eighty references are included.

APPROVED

FOR PUBLIC RELEASE

Distribution limited to Australian Government Agencies only. Other requests for this document must be referred to the Scientific and Technical Information Branch, Department of Defence, Canberra.

POSTAL ADDRESS: Assistant Secretary, Scientific and
Technical Information Branch,
Building 1, Room 2-14
Campbell Park Offices
CANBERRA ACT 2600

UNCLASSIFIED

409795

1B

CONTENTS

INTRODUCTION	1
USER NEEDS	2
Type of Information Need	3
Frequency and Turnaround	3
BACKUP	3
Periodicals	4
Non-Journal Material	4
STORAGE	5
Communications Services	5
Configuration	7
HARDWARE-SECONDARY STORAGE	8
SOFTWARE	8
FILES	9
RECORD STRUCTURE AND STATISTICS	10
SUMMARY	11
RECOMMENDATIONS	12
User Needs	12
System Design	13
Backup	13
Communications	14
Configuration	15
Hardware-Secondary Storage	15
Software	15
Records	16
APPENDIX I - Bradford's Law - graph	17
II - Relative effectiveness of different modes of retrieval	18
III- BLIPS record structure	19
IV - Satellite lease costs	20
REFERENCES	22

ACCESSION for

NTIS	White Section	<input checked="" type="checkbox"/>
DDC	Buff Section	<input type="checkbox"/>
UNANNOUNCED		<input type="checkbox"/>

JUSTIFICATION

BY

DISTRIBUTION/AVAILABILITY CODES

Dist.	AVAIL. NO. or SPECIAL
A	

STORAGE, RETRIEVAL AND DISSEMINATION OF INFORMATION:
CURRENT DEVELOPMENTS

INTRODUCTION

1. I was given the task of looking at current developments in storage, retrieval and dissemination of information.

2. It seemed rather pointless to reiterate procedures already known or followed, so I concentrated on finding alternatives. There were three main problems associated with this approach:

a. Apart from a rapid increase in popularity for on-line systems, there do not seem to have been any striking innovations over the last five years. So far as I can see, manual and mechanical methods have remained unchanged. (41,73) There is still one loud voice advocating manual SDI, a man called Davison who heads the Scientific Documentation Centre, and English commercial SDI service. A comparative evaluation was to be made between this Centre and an apparently badly executed computerised alternative named UKCIS (United Kingdom Chemical Information Service) by the Experimental Information Unit at Oxford, and the latter was going to be judged superior. However, this analysis fell into disrepute when Davison discovered that:

- (1) Both UKCIS and this experiment were funded by OSTI;
- (2) OSTI funds most of the Oxford Unit Experiments;
- (3) Certain variables in UKCIS performance which suggested a poor recall capacity were being suppressed. (8)

No further information analysing the relative merits of manual and computerised methods has been retrieved.

Mechanical indexing and retrieval methods such as optical coincidence do not seem to be any more popular or less cumbersome than 10-15 years ago. In some cases where mechanical methods have been applied post coordinately, there has been a relatively easy transition to computerised systems. (51,74)

So far as computer based systems are concerned, although there have been remarkable increases in hardware capacity, speed and reliability and decreases in size and cost, SRDI software has barely moved. It seems to have reached a zero sum situation, where an apparent advantage in a system is almost inevitably offset by a disadvantage. All my research has really dredged up is variations on a theme.

- b. The second problem follows from the first. Vast amounts of material had to be waded through to retrieve the finer details, so:

- (1) This paper does not claim to have covered all recent developments; and
- (2) I have not been able to sight anything like the range of material available; eg, proceedings of recent conferences, especially the annual ASIS conferences, and technical papers and the like, most notably those produced by Defence Documentation centres.

With the exception of a few not very recent conferences, my search has been confined mainly to journal articles, as these provided the most current material I could access.

- c. The third problem is self revealing: my overriding need was for currency, but I was caught by the problem of consulting rather elderly manual indexes, presumably surface mailed, to retrieve even more ancient references. Search strategy finally aimed at locating the most pertinent titles via the kind of information they contain, and hunting the shelves for the most recent issues. Problems b. and c. have meant that the information retrieved is uneven and this must inevitably affect this report.

3. In conclusion, it seems that the innovation stage for computerised SRDI systems is over, and they are now going through a period of consolidation. Perhaps the most useful literature at present is that which analyses experience.

USER NEEDS

4. Although the Defence sector response to the STISEC inquiries comprised a very small proportion of the total, it is assumed that the findings as a whole would reflect (to an extent) the heterogeneity which exists in this Department.

Type of Information Need

5. In descending order of frequency, it would appear that information needs are met by personal contacts (but not personal correspondence) scientific and technical journals, own organisations reports, trade journals and pamphlets, with textbooks and monographs, government reports and handbooks coming in about even. (64,ppl40-51) Of course, there are numerous variables affecting these figures: eg, dependency on Conference attendance may be quite high, but if appropriate conferences are not held very often, then frequency measurements are not representative. One interesting feature was the lack of interest in current awareness bulletins where they were available. Only the Australian Biochemical Society responses stood out. (64,pl42)

Frequency and Turnaround

6. If the STISEC findings are converted to averages, it appears that about 11% of users of STI regularly require information weekly, and 50% monthly. 11% require their information within an hour, 33% within a day and over 40% within a week. More than 50% felt that computer data bases would not be helpful, but as around 80% had had no formal training in information retrieval this may be as a result of ignorance. (64,ppl38-9) Two considerations here are that:

- a. This refers to the information itself, not agents to its retrieval. Interest in abstract journal, review journals and as mentioned, current awareness bulletins, tended to be fairly low.
- b. This Department's experience with ADSATIS, which may result in a different user perspective.

Apart from the STISEC report, I have not retrieved any information on methods for estimating information needs.

7. So far as outside users are concerned, both ANL and UNSW have expressed interest in using any system we put up.

BACKUP

8. I have not been able to retrieve anything very coherent or comprehensive. This is really the field of an experienced librarian, but three major problems seem to present themselves.

- a. The performance associated with retrieving classified material. This problem is going to occur no matter what system is introduced, and the only conclusion I can draw is the more rapid the retrieval system the more quickly requests can be put in the pipeline.

(1) Periodicals.

There are innumerable variations propounded on Bradford's law of scatter. This is normally applied to analysing subscription requirements. The method is to 'sum the references to each journal, rank the journals in order of decreasing productivity, and compute the cumulative references for each rank'. (56,p207)

In one review-come-test article I read the ranking was retrieved from the whole of one serially published bibliographic aid. Presumably this method could be converted to group profile retrievals - it should be reasonably easy to obtain statistics of journal titles cited.

A graph was created associating the cumulative number of articles with the ranked order. (See Appendix I). There is a tendency for an initial exponential increase followed by a general linear path, which tapers off at the end. The area between the origin and the beginning of the linear path is deemed to be the core collection. Selection on the linear path is a matter of library discretion.

Difficulties associated with this method are:

- Changes in journal name;
- References to abstracts and reviews reported in journals;
- that one title may produce a lot of chatty little articles, whereas another may have only three serious treatise. A way around this is to check the number of references per paper, as these are considered to be an indication of quality. COMPENDEX tapes record the number of references cited at the end of each abstract.

(2) Non-journal material.

It is possible that Bradford's law may be applicable to specialist sources - I have not heard of this being done. Presumably there are standing orders and gift & exchange deals with certain bodies in existence.

I retrieved one article which described an incredibly complicated and rather specious method for determining ordering priorities, but it did not assist with advance predictions. (21,15)

Probably the biggest problem with currency will be our proximity to the suppliers of documents. Current awareness lists will often be distributed well before the documents arrive. They even have this problem in England with US material, and it does arouse some user antagonism. (73)

- c. The third problem, as I see it, is the present extraordinary policy on photocopying. We must have the only library in the world which is required to distribute whole journals on inter-library loan. The excuse given is 'paper'. I am not sure what this signifies. If it is cost it is unreal when measured against the cost of multiple orders and the inevitable expense of repairing broken sets - especially as xerox works on economies of scale, which used to level off at 2 cents a sheet.

Apart from the direct costs are the indirect costs caused by extreme inconvenience eg, if 25 people want to sight one issue of a journal retrieved via a current awareness service, it could be a year before the last person gets it. If a journal is lost en route it takes a long time and a lot of money to get a reprint: if it is replaceable. Thirdly, if a user wishes to retain one article he must deny everyone, internal or external, access to the whole issue.

STORAGE

- 9. Developments associated with networking have led to concentration in the areas of transmission facilities, complex system integration and data base modelling.

Communications Services

- 10. Some perspective on Australia's facilities can be gained by looking at those available elsewhere. The US, Britain, France, Spain and shortly Canada all have packet switching facilities to coordinate data between transmission and reception points. These 'value Added' services include error detection and correction, automatic terminal type and speed recognition, data load control and facility to bypass faulty lines. (57,19,48, 47,42,37)

- 11. A battle is raging in the US between suppliers of lines. The giant common carrier, AT&T is conducting a price war with the specialised carriers, such as Datran and MCI. Unfortunately their method is to reduce charges in the heavily populated areas where the specialised carriers operate and to cover their losses by bumping up the charges to the less populous areas. The federal Communications Commission has stepped in to disallow this practice. (24,25,26)

12. The US, Canada and now Indonesia have domestic satellites, and Japan is about to put two up. India has one on loan from the US, but this is about to be reclaimed. The US has a surplus of satellites almost identical in type, launched by competing carriers. As they all operate on the same band, there is mutual interference which renders them unpopular. NASA recently launched a more versatile but limited life one for the Public Service Satellite Consortium, but it too is affected by microwave interference in the densely populated areas. It sounds as though the Indian 'gift' was an excuse to use some clean air for experiments: (maybe we could make NASA an offer?) India is now expected to 'develop a long term project of its own, dependent on national budgetary constraints'. (34,p93,26,48,42)

13. A communications technology which is gaining in popularity, especially in Japan, is fibre optics. This normally (but not necessarily) uses a laser source, light emitting diode transmitting agents and an optical cable receiver. Most recent developments are optical transistor amplifiers and reliable cables, eg. Corguide by Corning. This method exploits the extremely dense waveband capacity of light. Still needed are longer lived lasers. Signals can also be collected by a frequency modulator and transmitted to a LED and thus to the cable. This has been found to be very useful where space is at a premium. (47,48,78,pp39-40) Research on fibre optics is being carried out in Australia.

14. Australia should be putting up a satellite in the 1980's. (63)

15. According to a late 1974 document, services currently available in Australia on which data transmission can be effected are:

- a. Telex, which is slow, but cheap, uniform, compatible and reliable.
- b. Datel, which is a telephone network. Telecom has installed modems to facilitate networking on switched telephone lines. However, while it is considerably faster than Telex, it is still rather limited in the range of speeds available. It is expensive, charges being the same as for ordinary calls, and its analogue transmission gives rise to compatibility problems.
- c. Leased Lines, are very expensive, but have more capacity than telephone lines. Ability to transmit larger amounts of data via a much wider range of speeds tends to compensate for cost.
- d. CUDN, is the proposed digital network with inbuilt switching and transmission services. Apart from being cheaper to use than telephone or leased lines, this should save on terminal equipment and processor costs to the user. However, installation is well behind schedule. The National Library BIBDATA team was recently advised 'not to base any long term plans on the CUDN network'. (78,p7,55)

16. According to Dagmar Schmidmaier in 1975, (based on the Vernon Report) international telecommunications developments are outstripping internal developments, and it is on the cards that it will be cheaper to interrogate international data bases than national ones, all things considered. For further information on Library networking in Australia and overseas see: (63,77,60,59,45,44,40,35,20,12)

Configuration

17. Although the commonest configuration appears to be the star network, problems with giant computers are leading to more distributed configurations. Large computers pose extensive maintenance and reliability problems. Unreliability can lead to security difficulties, eg, the system can collapse through programming bugs. It is also claimed in a 1975 document that Large Scale Integrated Circuits warm up and give faults - however, I doubt that this would be repeated in 1976. (24,25,26) LSI Technology is now the latest and the greatest. An obvious result of a centralised breakdown is that all access points are affected simultaneously. (17,16)

18. Distribution may be by means of interconnected processors to work on segments. An extreme form of this is the INTEL Hypercube IV, which is 512 microprocessors arranged in a bank. It is not really successful, at present, because an error in one throws them all out. Computer users are beginning to be encouraged not to throw their old computers out when new fads come in, but to interconnect the new with the old. So far as I can see, Burroughs is the only company which has acted on this advice. (33,34)

19. ARPANET (Advanced Research Project Agency Network) is the best example of a fully distributed network, interlinking "dozens of computer installations". (26,p49) Communication is handled by modified minicomputers (Interface Message Processors) which perform packet switching functions. The commercial packet switching companies derived their technology from this development.

20. The star networks, such as Tymeshare, have devolved responsibility to satellite minicomputers, using a single file centre and the CPU as system controller. This controller also monitors the lines, which is rather different from the usual packet switching techniques. (49)

21. The next stages, in decreasing order from distributed to standard star networking include concentrators, which transmit selectively and at high speed formatted data to the CPU. These are composed of minicomputers, transmission control units or programmable multiplexors.

22. Ordinary multiplexors which take low speed input from a number of terminals and transmit it a high speed to the CPU, where it is again converted by a second multiplexor to conform to the computers intake capacity.

23. In its simplest form computer-terminal operations are facilitated by modems which take digital data, transform it to analog for higher speed transmission, and reconstruct it at the destination. (49,5,4,75,65,54,39,36,57,19)

HARDWARE - SECONDARY STORAGE

24. A revolutionary sounding, but surprisingly under-advertised device which is of great interest to those of us who want to store large numbers of large files has been developed by IBM and Control Data Corporation.

First introduced about a year ago, the IBM 3850 system combines wide-tape and large disk technologies to provide storage space for 30 to 236 billion bytes. This combination gives the fast access time and random retrieval capability of disks with the low price of tape. Each 3850 tape cartridge has a capacity of 50 megabytes. The cartridges are stored in a honeycomb nest that not only serves as a receptacle, but also provides physical security. Control Data Corporation's mass storage system, the OIC 3850, provides a minimum of 16 billion bytes of data that may be placed under the control of up to four IBM system 370 computers. (25,p47)

25. The largest disk (as far as I know) is the IBM 3336 Model 11 which stores over 200 megabytes. (49)

26. There are various small scale, self contained and portable devices currently being marketed. A popular example is the floppy disk, or diskette, with around 2 to 4 million bit capacity (250 - 500K bytes). (49,25) It is tending to be superseded by the wide tape cartridge programmable data entry machine, developed by Olivetti, Hewlett Packard, Tekronix and now IBM. Capacity is around 64K bytes. (80) These units maximise flexibility of access, are far cheaper to operate than large scale direct access devices, and would seem to have great potential for library housekeeping routines or report writing.

27. Other storage systems which are still very much in the developmental stage and seem to be considered more as main memory auxiliaries than direct access devices are the magnetic bubble and the semi conductor memories, especially the Charge Coupled Device. (2,24)

SOFTWARE

28. Software has tended to lag behind the demands created by increasing complex multi modal systems. (24,25,26)

29. Because of this there has been intense concentration on standardising software relationships between systems, (called software engineering) leading to the development of extensible languages, which can be readily altered and moved from one machine to another. (25,p49) This is mainly for minis and

universal compilers. A new area of interest is the hard software chip. The user would obtain the "chip module", a micro-computer type construction, from a catalogue of specifications, installing the modules he requires. This would eliminate the need for the various computer instructional programs which accompany software installation and represent more than 50% of programming needs. The user simply calls up the modules he requires, and the computer becomes an agent to predetermined functions without having to be instructed. (34) This procedure would be well adapted to repetitive tasks such as large file installation, retrieval and analysis.

30. Another software development pertinent to the information retrieval scene has been in data base modelling. Rather than rely on the usual slow space consuming methods of access (eg, indexes) access is determined by the properties of the data itself. Current developments are generally based on relationships between data elements, but the main inhibitor at present is a thorough prior knowledge of these relationships. (34,52,28,23,11)

FILES

31. That we should look to storing whole abstracts on direct access devices seems to be justified. There have been hundreds of retrieval tests performed using various techniques, some of which are so complicated with their weighting and nesting, and what have you, that it must take a very passive genius to cope on a day-to-day basis. The objective is invariably to avoid the expense of abstract storage. (72,61,53, 31,18,14,13,6,5 - Moureau).

32. One test which seemed worth reproducing was carried out in the US on COMPENDEX tapes, using 9 combinations of modes. (14) Fifty profiles were run, and it was assumed that the combination of all four modes (Title, Abstract, Subject, Free language) effected 100% retrieval. (See Appendix II). My criticisms of this test are:

- a. It does not exploit the advantages of interactive systems. Admittedly, this was beyond their scope.
- b. The profiles were not much good for the most part, consisting of an average of about 5 terms. If they were pre specified concepts the relationships should have been shown, but I doubt that they were. One example concerned computer design of computers, and the profile for it was computer-aided and computers with no truncation. Needless to say it yielded nothing from the abstract combinations - I also found few inexplicable inconsistencies that I won't go into now. (I also operated this profile on NTIS and got nothing - although I subsequently found a few references using a different search strategy).
- c. I have heard informally that COMPENDEX subject indexing is notoriously poor.
- d. There were no tests for precision, which, however subjective, seems necessary. (18,41,47,46)

33. Anomales noted by the authors were:

- a. That some items were not retrieved at all by title and abstract or abstract: most notably Information storage and retrieval. These terms only appeared in 27.4% of 285 apparently pertinent items.
- b. Abstract retrieval performed consistently but never outstandingly.

33. Finally, it was found that in comparison with all other combinations, abstracts combinations increased access time by a factor of 4, search time by a factor of 8 and file size by a factor of 10.

34. Most overseas organisations have resigned themselves to title, subject and free term searching, and there are numerous theories on computerised text searching for free terms to assist them. Usually this is a matter of discerning the ratio of terms to the size of the article, and reserving a prespecified proportion. There are very obvious problems with this approach, ie core terms may be rarely mentioned or, as shown, not at all. (662,32,30,29) Another method is for a clerk to indicate terms on the screen for automatic insertion in the index. DDC and NTIS tapes come supplied with a large range of uncontrolled as well as controlled descriptors.

RECORD STRUCTURE AND STATISTICS

35. Many publications supply detailed descriptions of record structures, including sample applications of logical and contextual retrieval operations.

36. One system which represents a wide range of input and output functions is that proposed by the British Department of the Environment Building Research Establishment (BLIPS). A possible criticism of this system is an apparent lack of concern with effecting compatibility either with externally available tapes or with outside systems, even though there is already an outside clientele from their old system (LIBINDEX). There was no suggestion, for instance, of consultation with EUSIDIC, the European standards organisation set up to facilitate networking. They are writing their own programs for ICL equipment. (51) An example of a system which has planned carefully for universality is the French Petroleum Institute. (5, pp107-18)

37. All the record is put onto disk initially online from a VDU terminal (see Appendix III). During overnight runs the following are produced:

- a. accessions lists;
- b. Abstracts bulletins in class order;
- c. Author and KWIK indexes;
- d. SDI profile matches.

38. These are copied onto a holding file if necessary.
Also:

- a. Loan records;
- b. Item cards, which are organised sequentially by item number;
- c. List of tags for editing and amendment.

Everything is then consigned to tape, with only the starred part retained on disk.

39. Flexibility in online retrospective operation is vested in the tags, to which controlled and uncontrolled terms are assigned at random. The index of terms is to be periodically checked and rationalised: otherwise profiling could be a nightmare.

40. Useful seeming are the possibilities of the loan records. Not only do they show who has what where and when, but they are excellent for storing statics of usage. Files can be purged over a period of years of unused material, thereby providing more space. This may be more accurately ascertained by storing SDI and retrosearch retrieval figures as well. I am sure that TYMESHARE must do something like this with their NTIS and ERIC tapes. Even allowing for the lower publications rate of earlier years, their files do not retain anything like the number of records that would have been produced in the time parameters they specify. (NTIS: 400,000 since 1964; ERIC: 140,000 since 1966). (9)

SUMMARY

41. This paper does not attempt to cover all areas of development in information storage, retrieval and dissemination. The primary objective, within specified constraints, was to look at developments not already familiar to the Scientific and Technical Information Branch. Directions to references are not only part of normal referencing procedure, but are also intended to fill out areas not expanded on.

42. Topics covered include:

- a. An overview of non-computerised systems.
- b. An examination of scientific and technical information requirements in Australia, derived from STISEC statistics.
- c. Problems and possible solutions in supplying backup, in particular those relating to currency. Aspects discussed included classified and non-classified foreign material, and the present policy on photocopying.

- d. Communications services: the Australian situation versus developments overseas, with emphasis on the United States, current technological innovations still in the development stage, and speculation on the relative merits of overseas and local data base interrogation.
- e. Types of hardware configuration in use overseas; developments, advantages and disadvantages: including star, partly distributed and fully distributed networks and the agents to their function.
- f. Recent secondary storage device development, including the small scale portable direct access device.
- g. Current software developments; the hard software chip, data base modelling and software engineering.
- h. The benefits and disadvantages of different modes of retrieval and methods for creating descriptors for tape-based files.
- i. An example of a multipurpose record structure and its operation, including a method for facilitating the purging of redundant files.

RECOMMENDATIONS

1. User Needs.

- a. That a user needs survey be conducted:
 - (1) to analyse types of user needs;
 - (2) to accurately estimate the user population.

See also 2.a, 3.a, 4.a, 1.c, and d.
- b. It is suggested that this be effected via a questionnaire, and that points to consider in this questionnaire may include an evaluation of:
 - (1) User knowledge of existing services;
 - (2) User suggestions on the improvement of existing services;
 - (3) User ranking of priorities in sources of information, based on value, not frequency of use.
 - (4) Information frequency and turnaround requirements.

(5) User identity of subject interests.

(6) Location of user.

It may be necessary to engage an advisor or consultant with professional qualifications and experience in questionnaire design and analysis; insofar as security permits.

- c. That predictions on future needs be extrapolated from the results of the questionnaire.
- d. It is suggested that no mention should be made of the possibilities of computerised services. Their existing and potential role should be able to be derived from statements on need and criticisms.

If it is decided to install a localised system (see 3a)

- e. That an ideal total system be planned according to the results and predictions stemming from the questionnaire. This ideal system may allow scope for modular introduction.
- f. That those in charge of Registry development be consulted.
- g. That an outside user needs analysis (of a more general nature) be considered if it is found that client expansion will achieve the following:
 - (1) Exploitation of the full potential of the system;
 - (2) A reduction in costs of system operation and maintenance as a result of:
 - economies of scale;
 - defrayal of costs by outside users;
 - backup distribution.
 - (3) A reciprocal cooperation with other library and information systems (in liaison with ALBIS) and the accompanying mutual benefits. (See also Section 4).
- h. That the questionnaire may be used to analyse gaps in user knowledge of services (1.b (1) and (2)) and that the best methods for rectifying this be ascertained.

2. Backup

- a. That the results of the questionnaire be used in discerning backup needs in advance of system implementation. (See especially Sections a(1), (2), b(2), (4), c.).
- b. That methods for effectively calculating backup requirements be investigated.

- c. That this Department might consider:
 - (1) large scale microform collection, for example, the Department may offer to relieve ANSTEL of their responsibility for certain NTIS COSATI categories and the retrospective collection.
 - (2) large scale microform creation, and in particular that Registry be consulted on the feasibility of converting old files to fiche. This may be modified by the Department's relationship with DIAC regarding reproduction of internal technical reports.
- d. That users be advised regarding currency problems involving:
 - (1) The acquisition of classified material from overseas. They should be informed on the necessary clearances and procedures and thus the resulting delays.
 - (2) The delays inherent in receiving foreign publications, caused by the economic necessity to use surface rather than airmail. That some heavily used serial publications or microforms be sent by airmail may be considered.
- e. That this Department cooperate in the development of BIBDATA, to facilitate and rationalise cataloguing, accessions, loans and interlibrary cooperation. It may be considered necessary to develop an independent but compatible system designed to hook readily into the BIBDATA system when it is installed.
- f. That the present policy on photocopying be altered where it concerns journal material, particularly in the case of:
 - (1) interlibrary loans;
 - (2) provision of current awareness and retrospective backup;
 - (3) The retention of articles by individuals.

3. Communications

- a. That a cost/benefit analysis be made comparing OTC (Telecom) links to foreign data bases with creating a local system;
 - (1) with reference to user population and location.

- b. That the availability and security of lines be analysed:

- (1) intradepartmental;
- (2) Telecom; It is suggested that the Department negotiate system design directly with Telecom planners.
- (3) Telecom projections (real).

See Appendix IV for costs involved in acquiring a satellite.

- c. That in the absence of packet switching services and digital lines, some method for the safe routing of data be established.

4. Configuration

- a. That an optimum configuration could be assisted by a user location analysis, derived from the user needs questionnaire.

- (1) to determine the most adequate centres for terminals;
- (2) to determine the number of terminals per centre;
- (3) following (1) and (2), to estimate processing capacity requirements.

- b. That overseas experience be evaluated when designing the network structure.

5. Hardware - secondary storage

- a. That an evaluation of the IBM 3850 or CDC 38500 mass storage systems be carried out:

- (1) taking into account the relative merits of abstracts versus non-abstracts retrieval.
- (2) In consultation with Registry in relation to online storage, retrieval and dissemination of files.

- b. That the potential of the small scale, portable direct access device be examined.

6. Software

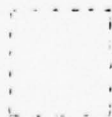
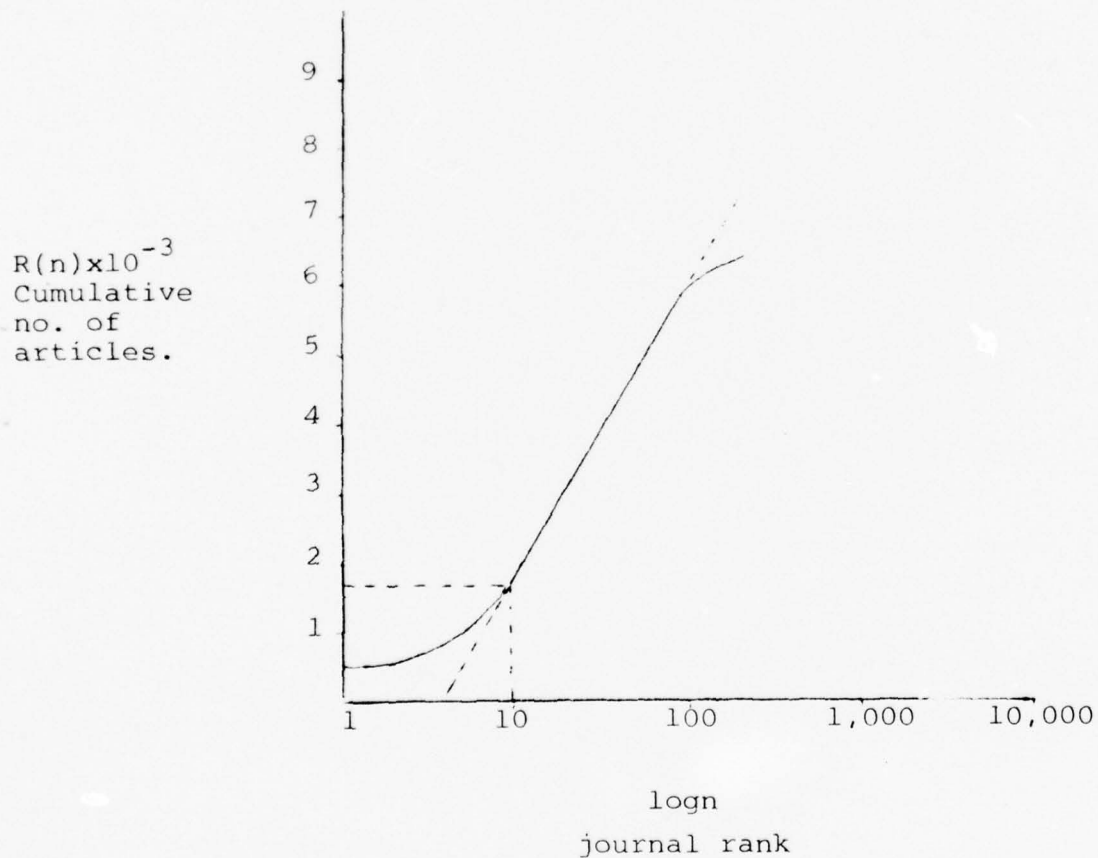
- a. That software developments which improve performance and save on installation, maintenance and storage costs be investigated, also those which involve intermachine compatibility.

7.

Records

- a. That some method for gathering statistics on pertinence and usage be collected so that files can be periodically purged of valueless items.

APPENDIX I



suggested core

Compiled from: Pope, A. Bradford's law and the periodical literature of information science. *Journal of the American Society for Information Science*, V 26 (4), July-August 1976:209.

APPENDIX II

RELATIVE EFFECTIVENESS OF TITLES, ABSTRACTS AND
SUBJECT HEADINGS FOR MACHINE RETRIEVAL FROM THE
COMPENDEX SERVICES

Title, Free Language, Subject, Abstract	100%
Title, Abstract	75%
Abstract	61%
Title, Free Language, Subject	47%
Title, Subject	41%
Subject, Free Language	30%
Title, Free Language	27%
Title	22%
Subject	21%

Compiled from: ASIS, V 26 (4), July-August 1975:
223-229. (Jerry R. Byrne)

APPENDIX III

BLIPS RECORD STRUCTURE

1. *Item Number
2. *Confidential Status
3. *Location
4. *Brief description
5. Full Bibliographic description
(5 subfields)
6. Abstract
7. Library Comments
8. *Tags
9. KWIC title
10. Sorting code for publications
11. *Loan data

Source: Neville H.H. and P.J. Elvin.

BLIPS: the Building Research Establishment Library processing system.
Aslb Proceedings, V 27 (5) May 1975:193.

APPENDIX IV

SATELLITE LEASE COSTS

The section below is reprinted from an article on the Public Service Satellite Consortium, written by members of the PSSC and the Communication Satellite Planning Center. (1)

It itemises the anticipated cost of setting up a fairly standard satellite communications system.

The capital cost of a satellite system consisting of two in-orbit satellites, a ground spare, and insurance to cover the risk incurred from launch through to in-orbit checkout is estimated to be \$(US) 89.4 million. For purposes of discussion, body-stabilised satellites similar to versions developed by General Electric and RCA are assumed. A nonrecurring charge of \$10 million is assumed to cover the modifications required by the Consortium for three higher-powered transponders on each spacecraft.

The cost breakdown is as follows:

Three satellites	\$36.0M
Two launches (Thor-Delta 3914)	28.0M
Hardware for possible third launch	9.0M
Development cost	10.0M
Subtotal	\$83.0M
Insurance for two launches	6.4M
Total	\$89.4M

The authors have learned from private conversations with NASA headquarters that the launch cost of a Thor-Delta 3914 (\$14.0 million) is the sum of \$9.0 million to the supplier of the vehicle and \$5.0 million to NASA for launch services. The hardware for a third launch (but not launch services) must be purchased in advance in order to be able to react quickly in the event of catastrophe.

The cost of insurance is estimated to be 10 percent of the value of the in-orbit investment, exclusive of non-recurring costs. This rule-of-thumb is commonly used in the industry.

Annual revenue requirements are calculated on the basis of a return on capital investment of 21 percent over seven years. The required payback factor is 28.5 percent, leading to required annual revenue of \$25.5 million. To this total is added the estimated marginal cost of telemetry, command, and control, which is assumed to be \$2.0 million annually.

The annual tariff paid by the Consortium for six in-orbit transponders (three on each satellite) is assumed to be based on its relative utilisation of satellite capacity. Sufficient power and weight are available to the communications subsystem on a General Electric satellite to use 17 20-watt transponders at the beginning of life and 12 20-watt transponders at the end of life. Spacecraft vendors typically place more transponders on a spacecraft than could be powered at the end of life, recognising that some fraction of the transponders may eventually fail.

To determine the proportion of the satellite capacity consumed if the Consortium wished to use only three 20-watt transponders on each satellite, one would have to know precisely how the remaining capacity was utilised. For concreteness, the tariff paid by the Consortium will be calculated under the assumption

that there are 14 identical 20-watt transponders on each spacecraft, or 28 transponders in orbit. Assume there is 50 percent utilisation of the 22 transponders not used by the Consortium. Then the Consortium would utilise 35.3 percent of the satellite capacity subscribed for, and should pay the carrier \$9.71 million annually for the satellite service.

Assume that the six transponders leased by the Consortium are utilised 40 percent of the time over the course of a year. Then the per-hour cost of transponder time would be approximately \$463. Assume that the Consortium applied a 20 percent margin to allow for underutilisation of transponder time by the membership. Then the price to a member of one hour of transponder time would be approximately \$556.

- (1) Lusignan, B.B., Potter, J.G. and Janky, J.M. The co-op in the sky. IEEE Spectrum, V13(2), Feb 1976:63.

REFERENCES

1. ALLAN, Roger
Circuit/System building blocks. IEEE Spectrum V12(1).
Jan 1975:49-53.
2. ALLAN, Roger
Components: microprocessors galore. IEEE Spectrum.
V13(1). Jan 1976:50-56.
3. ALLAN, Roger
Technology forecasting V: components. IEEE Spectrum.
V12(4). April 1975:56-9.
4. ALLERY, G.D.
Data communications and public networks. Information
Processing 74. Proceedings of IFIP Congress 74,
organized by the International Federation for Information
Processing, Stockholm, Sweden, August 5-10, 1974, ed. by
Jack L. Rosenfeld. Amsterdam, North-Holland Publishing
Company, 1974:117-22.
5. AMERICAN SOCIETY FOR INFORMATION SCIENCE. Proceedings
Vol 9.
A world of information. Proceedings of the American
Society for Information Science 35th Annual Meeting,
Washington DC, October 23-26, 1972. Vol 9. Edited by
John H. Wilson. Washington DC, ASIS, 1972.
Tocatlian, J. UNISLIST implementation plans, pp9-14.
TURKOV, A. The international nuclear information
system (INIS): practical and political
aspects of international co-operation in
information exchange. pp15-22.
ALLEN, T.J. and COONEY, S. National and international
communication networks and the transfer of
technology. pp25-6.
BLOCK, Uri. Is it technology that impedes information
network development? pp65-70.
REICHARDT, G. The use of available information systems by
special libraries. pp71-8.
ABORN, P.K. and others. Supplying machine readable
data bases to users outside the United
States.
MOUREAU, M. and others. Search strategies at the
Institut Francais du Petrole using
non-French services. pp107-18.
EDELSON, B. Satellite communications for information
networking. pp175-88.

- KALLENBACK, P.A. Towards a European information utility network. pp197-204.
6. ANGIONE, P.V.

On the equivalence of Boolean and weighted searching based on the convertibility of query forms. Journal of the American Society for Information Science, V26(2), March-April 1975:112-124.
 7. ANNUAL Review of Information Science and Technology, V8-1973-. Washington, DC, American Society for Information Science, 1973, Carlos A. Cuadra, ed.
 8. ANON

The Icarus file
Information Scientist V9(1) March 1975:44-
 9. ANSTEL NEWS.
 10. BLANKEN, R.R. and Stern, B.T.

Planning and design of online systems for the ultimate user of biomedical information. Information Processing and Management, V11(8/12) 1975:207-28.
 11. BOOKSTEIN, Abraham

Hash coding with a non-unique search key. Journal of the American Society for Information Science (ASIS) V25(4) July-August 1974:232-6.
 12. BUTLER, Brett

State of the nation in networking. Journal of Library Automation V8(3) Sept 1975:200-20
 13. BUTTERLY, Elizabeth

Improving SDI search profiles. Information Processing and Management, V11(8/12) 1975:189:200
 14. BYRNE, J.R.

Relative effectiveness of titles, abstracts and subject headings for matching retrieval from the COMPENDEX services. Journal of the American Society for Informations Science, V26(4) July-Aug 1975:223-9.
 15. CAWKELL, A.E.

A measure of 'Efficiency Factor' - communication theory applied to a document selection systems. Information Processing and Management, V1(8/12) 1975:243-8.

16. CHRISTIANSEN, D.
Applications review: technology '76. IEEE Spectrum, V13(1), Jan 1976:42-3.
17. CHRISTIANSEN, D.
Technology forecasting: Time and technology: looking ahead in electronics. IEEE Spectrum, V12(4), April 1975:41
18. COOK, Kenneth H.
A threshold model of relevance decisions. Information Processing and Management, V11(5/7) 1975:125-36
19. DAVIES, D.W.
Packet switching, message switching and future data communication networks. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1975:147-50.
20. DAVIES, G.W.P.
Information networks in the European community. The Information Scientist, V9(3) Sept 1975:91-8
21. DE PEW, J.N.
An acquisitions decision model for academic libraries, Journal of the American Society for Information Science, V26(4) July-Aug 1975:237-46.
22. ELVIN, P.J. and H.H. Neville
Extended use of the BLIPS system. Aslib Proceedings, V27(5) May 1975:210-16
23. ESWARAN, K.P.
Placement or records in a file and file allocation in a computer network. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:304-7
24. FALK, Howard
Computer systems: hardware/software. IEEE Spectrum, V12(1) January 1975:38-42.

25. FALK, Howard
Computers: poised for progress. IEEE Spectrum, V13(1), Jan 1976:44-9.
26. FALK, Howard
Technology forecasting 1: communications
Technology forecasting 11: computers
IEEE Spectrum, V12(4) April 1975:42-50
27. FRYCHI, S.J. and others
Squibb Science Information System: computerised selective dissemination, current awareness and retrospective searching of pharmaceutical literature. Journal of the American Society for Information Science, V26(3), May-June 1975: 174-83
28. CRONER, L.H. and A.L. Goel
Concurrency in hashed file access. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:431-35
29. HARTER, S.P.
A probabilistic approach to automatic keyword indexing, part 1. Journal of the American Society for Information Science, V26(4) July-Aug 1975:197-206
30. HARTER, S.P.
A probabilistic approach to automatic keyword indexing. Part 2. Journal of the American Society for Information Science. V26(5), Sept-Oct 1975:280-9
31. HUMPHREY, S.M.
Searching the MEDLARS citation file online using ELHILL and STAIRS: an updated comparison. Information Processing and Management, V12(1), 1976:63-70
32. HUNT, B.L. and others
Machine-assisted indexing of scientific research studies. Journal of the American Society for Information Science, V26(4), Jul-Aug 1974:230-6
33. IEEE SPECTRUM
Special report: technology '75. IEEE Spectrum, V12(1) Jan 1975

34. IEEE SPECTRUM
Technology '76. IEEE Spectrum, V13(1) Jan, 1976:42-94
35. JACAB, M.E.
Summary of Lasie Seminar on Library Networks and Cooperative efforts. Lasie, V5(3) Nov/Dec 1974:3-5
36. KAISER, C. and KRAKOWAIIK, S.
Design and implementation of a time-sharing system: a critical appraisal. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:247-51
37. KALLENBACH, P.A.
Introduction to data transmission for information retrieval. Information processing and Management, V11(5/7) 1975:137-46
38. KATTER, R.V. and Pearson, K.M.
MEDLARS 11: a third generation bibliographic production system. Journal of Library Automation, V8(2), June 1975:87-97
39. KLEINROCK, L.
Resource allocation in computer systems and computer-communication networks. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:11-18
40. KOUNTZ, J.C.
Library support through automation: the California State University and colleges plan for library automation. Journal of Library automation, V8(2) June 1975:98-114
41. LANCASTER, F.W.
Information retrieval systems: characteristics, testing and evaluation. New York, John Wiley, 1968
42. LUSIGNAN, B.B. and others
The coop in the sky. IEEE Spectrum, V13(2) Feb 1976:59-65

43. LYNCH, M.F.
Computer-based information services in science and technology-principles and techniques. Stevenage, Herts., Peter Peregrinus, 1974
44. MAGUIRE, Carmel
Library networks and the Australian scene part 1 - Introduction to networks
part 2 - The Australian scene, by Dorothy Peak
Lasie, V5(3), Nov-Dec 1974:4-17
45. MALTBY, G.R.
Lasie Library Network Seminar - Facilitied for cooperative efforts and networks, Lasie, V5(3) Nov/Dec 1974:28-39
46. MEADOW, C.T.
The analysis of information systems, 2 ed. Los Angeles, Calif., Melville Publishing, 1973. A Wiley-Becher & Hayes series book.
47. MENNIE, D.
Communications: electronics and optics too.
IEEE Spectrum, V13(1), Jan 1976:57-63
48. MENNIE, D.
Communications and microwave. IEEE Spectrum, V12(1), Jan 1975:43-48
49. MILLER, R.H., Bliel, J.M. and Corcoran, J.W.
Innovation in the dissemination of scientific and technical information. A management planning guide compiled for the National Science foundation Office of Science Information Service, Contract NSF-C951, April 1975. Stanford, Applied Communication Research, 1975, reproduced by National Technical Information Service, no PB 244 333.
50. MOGH DAM, D.
User training for on-line information retrieval systems. Journal of the American Society for Information Science, V26(3), May-June 1975:184-8
51. NEVILLE, H.H. and ELVIN, P.J.
BLIPS: the Research Establishment Library processing systems. Aslib Proceedings, V27(5) May 1975:189-203

52. NEWMAN, W.L. and COXFORD, F.H.
Blocking and deblocking variable length records in a direct access file. Journal of Library Automation, V8(2), June 1975:115-26
53. NICOLAS, J.M. and SYRE, J.C.
Natural language question answering and automatic deduction in the system SYNTEX. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:595-9
54. NIELSON, N.R.
The implications of star computing networks. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:908-12
55. O'HANLON, M.S.
British Library research and development: current projects and future trends. Information Scientist, V9(3) Sept 1975: 99-106
56. POPE, A.
Bradford's law and the periodical literature of information science. Journal of the American society for Information Science, V26(4), July-Aug 1975:207-13
57. POUZIN, L.
CIGALE, the packet switching machine of the CYCLADES computer network. Information Processing, Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974 ed by Jack L. Rosenfeld. Amsterdam, North-Holland, Publishing Company, 1974:155-9
58. PRANGNELL, R.D. AND P.J. ELVIN
Information retrieval by computer at the Building Research Establishment Library. The Information Scientist, V9(1), March 1975:3-18
59. REED, M.J.P.
The Washington Library Network's computerised bibliographic system. Journal of Library Automation, V8(3) Sept 1975:174-99

60. RICHARDSON, W.D.
The National Library of Australia's role in cooperative efforts and networks. Lasie, V5(3) Nov-Dec 1974:18-27
61. ROBERTSON, S.E.
Explicit and implicit variables in information retrieval systems. Journal of the American Society for Information Science, V26(4), July-Aug 1975:214-22
62. SALTON, G., WONG, A AND YU, C.T.
Automatic indexing using term discrimination and term precision measurements. Information Processing and Management, V12(1), 1976:43-52
63. SCHMIDMAIER, Dagmar
Machine readable data bases in Australia: a state of the art report. LASIE, V5(6), May-June 1975:2-74
64. SCIENTIFIC AND TECHNOLOGICAL INFORMATION SERVICES ENQUIRY COMMITTEE
The STISEC report. Report to the Council of the National Library of Australia by the Scientific and Technological Information Services Enquiry Committee: May 1973.
Canberra, National Library, 1973
Vol 1. Scientific and Technological information services in Australia.
Vol 2. Procedures, evidence examined, findings and appendixes.
65. SHARMA, R.L. and others
C-System: multiprocessor network architecture. Information Processing 74. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed. by Jack L. Rosenfeld. Amsterdam, North-Holland Publishing Company, 1974:19-23
66. STEINACKER, Ivo.
Indexing and automatic significance analysis. Journal of the American Society for Information Science, V25(4) July-Aug 1974:237-41
67. SWANSON, R.W.
Performing evaluation studies in information science. Journal of the American Society for Information Science, V26(3) May-June 1975:140-56

68. TAGLIACOZZO, R.

The consumers of new information technology: a survey of the utilisation of MEDLINE, Journal of the American Society for Information Science, V26(5) Sept-Oct 1975:294-304

69. TURN, R.

Computers in the 1980's - trends in hardware technology. Information Processing 74. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed. by Jack L. Rosenfeld. Amsterdam, North-Holland publishing Company, 1974:137-40

70. UNDY, Gordon

Summary of an overseas trip to investigate large scale computer systems encompassing total library operations. LASIE, V5(4) Jan-Feb 1975:13-20

71. VALLANCE, L.S.

Computer aspects of the BLIPS system. Aslib Proceedings, V27(5) May 1975:204-9

72. VAN RIJSBERGEN, C.J. and CROFT, W.B.

Document clustering: an evaluation of some experiments with the Cranfield 1400 collection. Information Processing and Management, V11(5/7) 1975:171-82

73. VINCENT, Ida and Seals, J.H.

A manual current awareness service at the University of Aston. Aslib Proceedings, V27(6), June 1975:247-261

74. WALL E.

Symbiotic development of Thesauri and information systems; a case history. Journal of the American Society for Information Science, V26(2), March-April 1975:71-93

75. WEDBURG, G.H. and HUASCHILE, L.W.

The General Electric network monitor system. Information Processing. Proceedings of IFIP Congress 74, organised by the International Federation for Information Processing, Stockholm, Sweden, August 5-10, 1974, ed by Jack L. Rosenfeld. Amsterdam, North-Holland publishing Company. 1974:24-8

76. WILDE, D.U. and CHAFE, H.D.

Reduction of data generation costs. Journal for the American Society for Information Science, V26(2), March-April 1975:103-7

77. WILLIAMS, M.E.
Progress and problems of the data base community.
Journal of the American Society for Information Science,
V26(5), Sept-Oct 1975:305-6
78. ELECTRONIC International, V49(11) May 27 1976
79. MALTBY G.F. Facilities for cooperative efforts and
networks
LASIE, V5(3) Nov-Dec 1974:28-29
80. ELECTRONICS WEEKLY, 5/5/76:3

DOCUMENT CONTROL DATA SHEET

Security classification of this page

-
- | | |
|-----------------------------------|--|
| 1. DOCUMENT NUMBERS | 2. SECURITY CLASSIFICATION |
| a. AR Number : AR000343 | a. Complete document:
unclassified |
| b. Document Series and Number: | b. Title in isolation:
unclassified |
| c. Report Number STIB REPORT 2/76 | c. Summary in Isolation:
unclassified |
-
3. TITLE:
- STORAGE RETRIEVAL AND DISSEMINATION OF INFORMATION:
CURRENT DEVELOPMENTS
-
- | | |
|-----------------------|-------------------|
| 4. PERSONAL AUTHOR(S) | 5. DOCUMENT DATE: |
| CLARKE, J. | JULY 1976 |
-
- | | |
|--|--|
| 7. ISSUING AUTHORITY(S): | 6. TYPE OF REPORT AND
PERIOD COVERED: |
| Scientific and Technical
Information Branch | N/A |
-
- | | |
|---------------------------|-------------------|
| 8. REFERENCE NUMBERS | 9. COST CODE: N/A |
| a. Task: N/A | |
| b. Sponsoring Agency: N/A | |
-
- | | |
|--|-----------------------------|
| 10. IMPRINT (Publishing establishment): | 11. COMPUTER PROGRAM(S) |
| Scientific and Technical
Information Branch | (Title(s) and languages(s)) |
| | N/A |
-
12. RELEASE LIMITATIONS (of the document);
- Nil
-
13. ANNOUNCEMENT LIMITATIONS (of the information on this page):
- Nil
-
- | | |
|--------------------------|------------------|
| 14. DESCRIPTORS. | 15 COSATI CODES: |
| Information retrieval | 0502 |
| Reviews | |
| Computer storage devices | |
| Networking | |
-
16. SUMMARY (if this is security classified, the announcement of this
OR report will be similarly classified):
ABSTRACT
- A literature survey of current developments in storage,
retrieval and dissemination of information and associated
technologies, with assessment of their application to the
Department of Defence. 80 refs.
-

DISTRIBUTION

DEFENCE CENTRAL

CDS

ECADSS

SDSA

PL

CAMPBELL PARK LIBRARY

WRE

LIBRARY

ARL

LIBRARY

MRL

LIBRARY